Example: Application of the GPC Model to a Reasoning Ability Instrument, MMLE, MULTILOG

As an example, we apply the GPC model to the Alike Reasoning data calibrated with the PC model in Chapter 7 and in MULTILOG_PCMcalibrationEx.pdf. Although PARSCALE could be used to perform this calibration, for comparison purposes with the PC model results we use MULTILOG. In addition, we demonstrate the analysis of pattern data with MULTILOG. Pattern data consist of all the unique patterns and their corresponding frequencies. (On the author's website is the output from the analysis of these data using the generalized partial credit model and PARSCALE.)

The command file is identical to that used for the PC calibration (MULTILOG_PCMcalibrationEx.pdf: Table 7.1) except for three changes. First, to use pattern data we make two changes to the `PROBLEM` line: (1) the keyword `PATTERNS` is used in lieu of `INDIVIDUAL`, and (2) `NPATTERNS = 6561` is specified instead of `NEXAMINEES=3000`. (Given the number of individuals, we do not observe all of the 3^8 = 6561 possible patterns; only 904 unique patterns are observed.) The third change is that the line '`EQUAL ALL AK=1`' used in the PC calibration is omitted to allow for the discrimination parameter estimates, $\hat{\alpha}_j$ s, to vary across items. The output file has the same appearance as that of the PC model calibration except for the additional output containing the EAP $\hat{\theta}$ for each pattern. That is, when MULTILOG calibrates pattern data it produces item and person parameter estimates in a single calibration run.

Table 1 contains the abridged output. As can be seen from the `ESTIMATION PARAMETERS` section, the `NUMBER OF FREE PARAMETERS` is `24` (i.e., each item has 3 parameters (2 transition locations, δ_{j1} and δ_{j2}, plus 1 α_j) times 8 items). The first line in the data file consists of a response vector of all zeros (i.e., `ITEMS 00000000`) and there are 53 persons (i.e., `WT/CR 53`) who provided this set of responses. Convergence is achieved in 16 iterations.

Table 1. Abridged output from MULTILOG GPC model calibration example.

```
    :
DATA PARAMETERS:
 NUMBER OF LINES IN THE DATA FILE: 6561
 NUMBER OF CATEGORICAL-RESPONSE ITEMS:   8
 NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS:   1
 TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS):   9
 NUMBER OF CHARACTERS IN ID FIELDS:  8
 MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM:  3
 THE MISSING VALUE CODE FOR CONTINUOUS DATA:  9.0000
 RESPONSE-PATTERN FREQUENCIES WILL BE READ
 THE DATA WILL BE STORED IN SCRATCH FILES ON DISK

ESTIMATION PARAMETERS:
 THE ITEMS WILL BE CALIBRATED--
   BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION
 MAXIMUM NUMBER OF EM CYCLES PERMITTED: 100
 NUMBER OF PARAMETER-SEGMENTS USED IS:   8
 NUMBER OF FREE PARAMETERS IS:   24
 MAXIMUM NUMBER OF M-STEP ITERATIONS IS  50 TIMES
   THE NUMBER OF PARAMETERS IN THE SEGMENT
 THE M-STEP CONVERGENCE CRITERION IS: 0.000100
 THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000
      :
KEY-
CODE  CATEGORY
 0     11111111
 1     22222222
 2     33333333
      :
FIRST OBSERVATION AS READ-
ID    00000000
ITEMS 00000000
NORML      0.000
WT/CR      53.00
      :
FINISHED CYCLE  16
MAXIMUM INTERCYCLE PARAMETER CHANGE=   0.00059 P(   2)
      :
ITEM   1:      3 NOMINAL CATEGORIES,  3 HIGH
 CATEGORY(K): 1      2      3
   A(K)     -1.30   0.00   1.30
   C(K)      0.00   2.00   4.24

          CONTRAST-COEFFICIENTS (STANDARD ERRORS)
 FOR:              A                     C
 CONTRAST P(#)  COEFF.[POLY.]  P(#)  COEFF.[ TRI.]
     1       1   1.30 (0.07)      2  -2.00 (0.16)
     2      25   0.00 (0.00)      3  -2.24 (0.09)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.327  0.433  0.566  0.722  0.887  1.037  1.139  1.168
-1.4 -  0.0  1.114  0.993  0.835  0.671  0.521  0.397  0.298  0.223
```

```
 0.2 -  1.6  0.166  0.125  0.093  0.070  0.053  0.040  0.031  0.023
 1.8 -  3.0  0.018  0.014  0.011  0.008  0.006  0.005  0.004

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3
 OBS. FREQ.    201    351   2390
 OBS. PROP.  0.0683 0.1193 0.8124
 EXP. PROP.  0.0668 0.1210 0.8122

   :

@THETA:      INFORMATION:
-3.0 - -1.6  1.777  1.970  2.206  2.486  2.798  3.118  3.411  3.646
-1.4 -  0.0  3.801  3.879  3.895  3.872  3.827  3.774  3.721  3.665
 0.2 -  1.6  3.601  3.520  3.416  3.286  3.132  2.959  2.776  2.593
 1.8 -  3.0  2.416  2.251  2.102  1.970  1.854  1.753  1.664

@THETA:      POSTERIOR STANDARD DEVIATION:
-3.0 - -1.6  0.750  0.713  0.673  0.634  0.598  0.566  0.541  0.524
-1.4 -  0.0  0.513  0.508  0.507  0.508  0.511  0.515  0.518  0.522
 0.2 -  1.6  0.527  0.533  0.541  0.552  0.565  0.581  0.600  0.621
 1.8 -  3.0  0.643  0.667  0.690  0.712  0.734  0.755  0.775

MARGINAL RELIABILITY:    0.7058

OBSERVED(EXPECTED)    STD.  :     EAP (S.D.)  :  PATTERN
                      RES.  :                 :

    53.0(    24.9)    5.62  :   -2.13 ( 0.60) :  11111111
     2.0(     1.9)    0.08  :   -1.91 ( 0.57) :  11111112
     0.0(     0.2)   -0.42  :   -1.70 ( 0.55) :  11111113
     0.0(     0.9)   -0.94  :   -1.84 ( 0.56) :  11111121
     0.0(     0.1)   -0.28  :   -1.63 ( 0.54) :  11111122
     0.0(     0.0)   -0.09  :   -1.44 ( 0.53) :  11111123
     0.0(     0.2)   -0.44  :   -1.57 ( 0.54) :  11111131
     0.0(     0.0)   -0.14  :   -1.38 ( 0.53) :  11111132
     0.0(     0.0)   -0.05  :   -1.20 ( 0.52) :  11111133
    12.0(     8.2)    1.31  :   -1.92 ( 0.57) :  11111211
      :
     8.0(     7.6)    0.16  :    2.11 ( 0.67) :  33333333

NEGATIVE TWICE THE LOGLIKELIHOOD=      3071.4
(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)
```

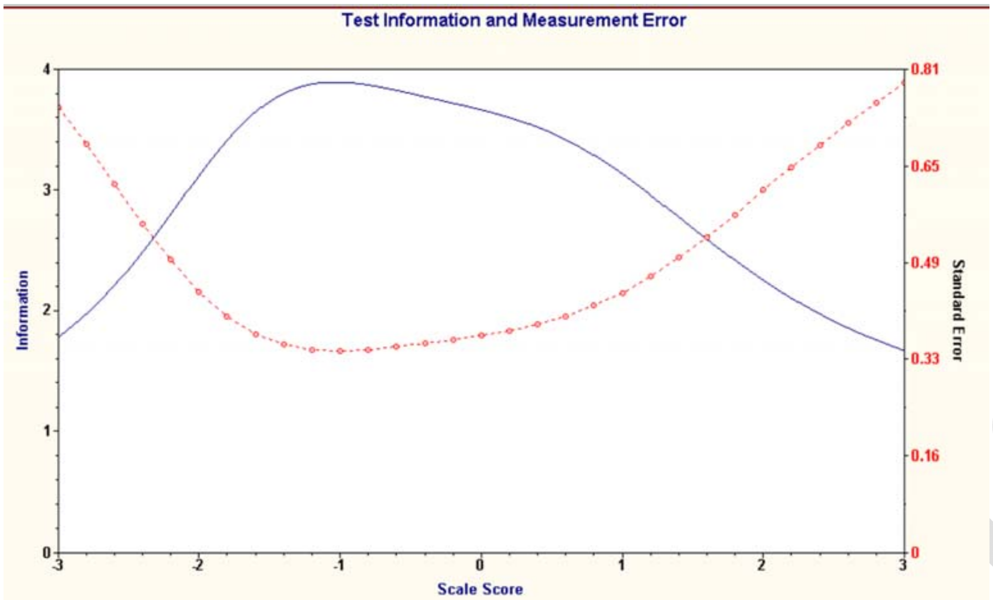
As done in MULTILOG_PCMcalibrationEx.pdf, the item parameter estimates are read from the `CONTRAST-COEFFICIENTS` table. The rest of the item parameter estimate output is similar to that seen with the PC model and is interpreted the same way. Table 2 contains the item parameter estimates and their corresponding standard errors for all the items on the instrument. The $\hat{\alpha}_j$ column shows that the items differ in their discrimination. By and large, the transition locations are found to fall between –2.24 and 2.42 and so it is not surprising that the total information function (Figure 1) indicates that this instrument tends to do a better job estimating individuals located within this range than outside of it. In addition, the GPC total information maximum is located around –1.2 and is more peaked than that observed with the PC model.

Table 2. Item parameter estimates from MULTILOG GPC model calibration example.

Item	$\hat{\alpha}_j$	$s_e(\hat{\alpha}_j)$	$\hat{\delta}_{j2}$	$s_e(\hat{\delta}_{j2})$	$\hat{\delta}_{j3}$	$s_e(\hat{\delta}_{j3})$
1	1.30	(0.07)	-2.00	(0.16)	-2.24	(0.09)
2	1.06	(0.06)	-0.64	(0.10)	-1.34	(0.07)
3	0.84	(0.05)	-1.04	(0.09)	-0.24	(0.05)
4	0.67	(0.04)	-0.22	(0.06)	0.51	(0.06)
5	0.87	(0.04)	1.57	(0.09)	-1.21	(0.08)
6	0.63	(0.05)	-0.16	(0.05)	2.42	(0.11)
7	0.88	(0.05)	1.60	(0.07)	0.03	(0.09)
8	0.66	(0.05)	1.24	(0.05)	1.18	(0.10)

MULTILOG provides –2ln*L* at the end of its output (i.e., `NEGATIVE TWICE THE LOGLIKELIHOOD`). As mentioned in Chapter 7, when pattern data are calibrated this index is positive and may be used for an overall assessment of model–data fit; this assumes that most or all of the possible patterns are observed (i.e., there are very few cells with zero frequencies in the contingency table made up of m^L cells). With pattern data the –2ln*L* value can be interpreted as a chi-square with *df* = [(number of patterns) – (number of estimated item parameters) – 1], where the number of estimated item parameters can be determined by the user or obtained from the line '`NUMBER OF FREE PARAMETERS IS`' and the (number of patterns) = m^L. For example, with four three-response category items, L = 4 and *m* = 3, the number of patterns is 3^4 = 81. The null hypothesis is that the data are consistent with a general multinomial model. In our example the combination of the number of items, the number of response categories, and the sample size result in a contingency table that is quite sparse (i.e., there are many cells with zero frequencies and small expected frequencies) and the –2ln*L* is not distributed as a chi-square.

Figure 1. Total information for Alike reasoning exam, GPC model.[a]



[a]Legend-Solid line: total information, Dotted line: Standard error.

Alternatively, as shown in Chapter 6, the difference in –2ln*L*s from two hierarchically nested models is distributed as a chi-square with *df* equal to the difference in the number of parameters between the Full model and the Reduced model, given that the Full model holds for the data. The calibration of these pattern data with the PC model produces a –2ln*L* of 3148.7 with 17 free parameters. Given that the PC model is nested within the GPC model, the difference between their respective –2ln*L*s is distributed as a chi-square with a value of 77.2 with 7 degrees of freedom; we assume that we would observe GPC model–data fit with a larger sample. Therefore, the GPC model fits these data significantly better than does the PC model. One reason for the GPC's better fit is its capability to capture the variability in the item discriminations. As mentioned above, as part of any model–data fit analysis one should compare the empirical and predicted ORFs. The program MODFIT could be used for this purpose with the GPC model.